

Appendix contents

Supplementary figures.....2

Supplementary methods.....5

References for supplementary methods.....7

Primers used in the study.....8

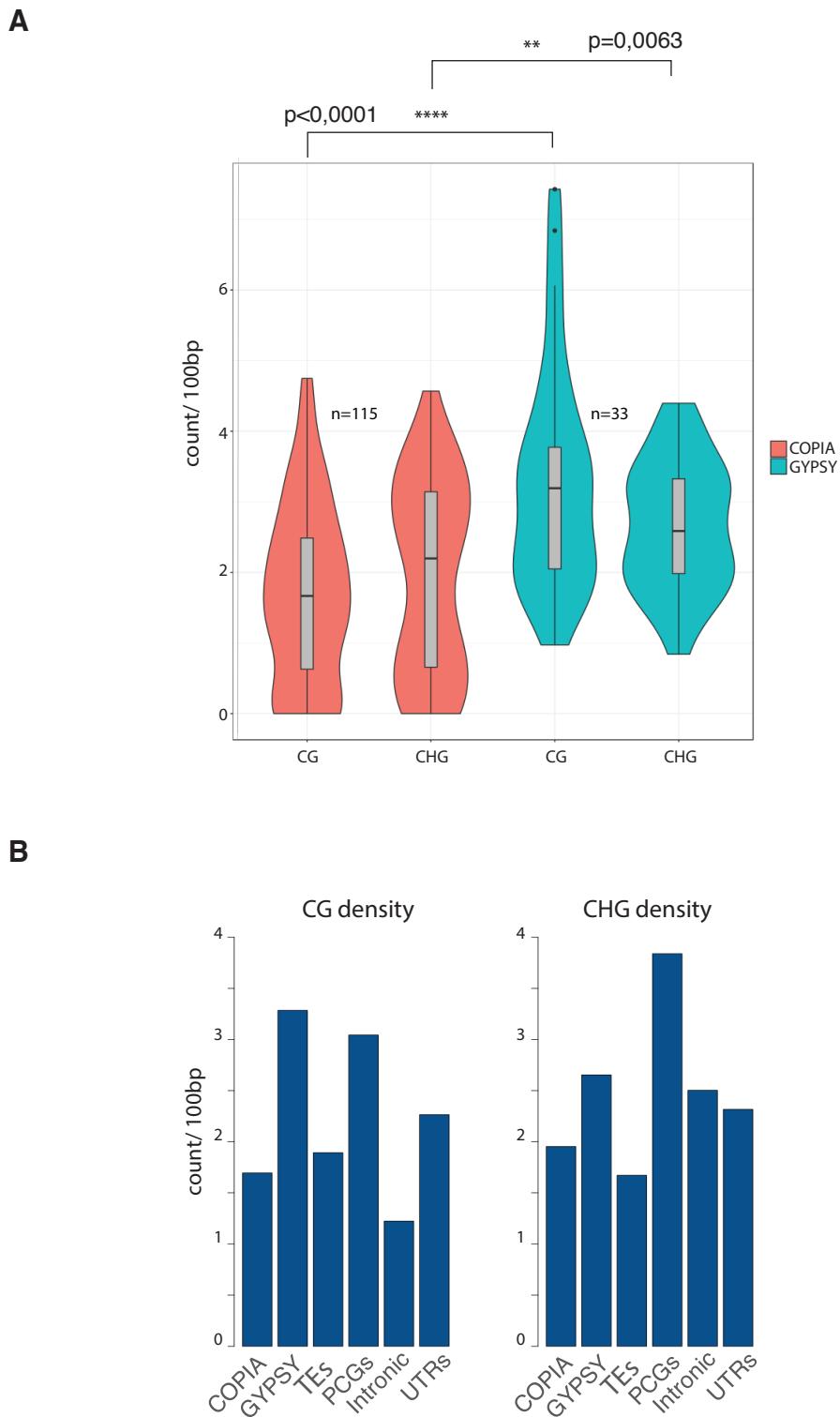


Figure S1. Genome-wide analysis of CG and CHG di- and tri-nucleotides frequencies at *Arabidopsis* COPIA elements.

A. CG and CHG density in number of CG dinucleotides and CHG trinucleotides counted per 100 bp (% CG) at COPIA and GYPSY LTRs (Rebase prototypes).

B. Average CG (left) and CHG (right) density at COPIA and GYPSY LTRs compared to the densities at all annotated segments in the *Arabidopsis thaliana* genome.

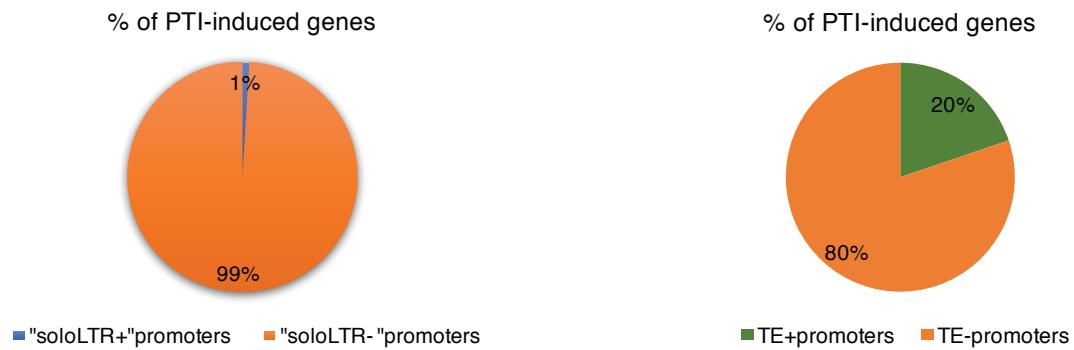
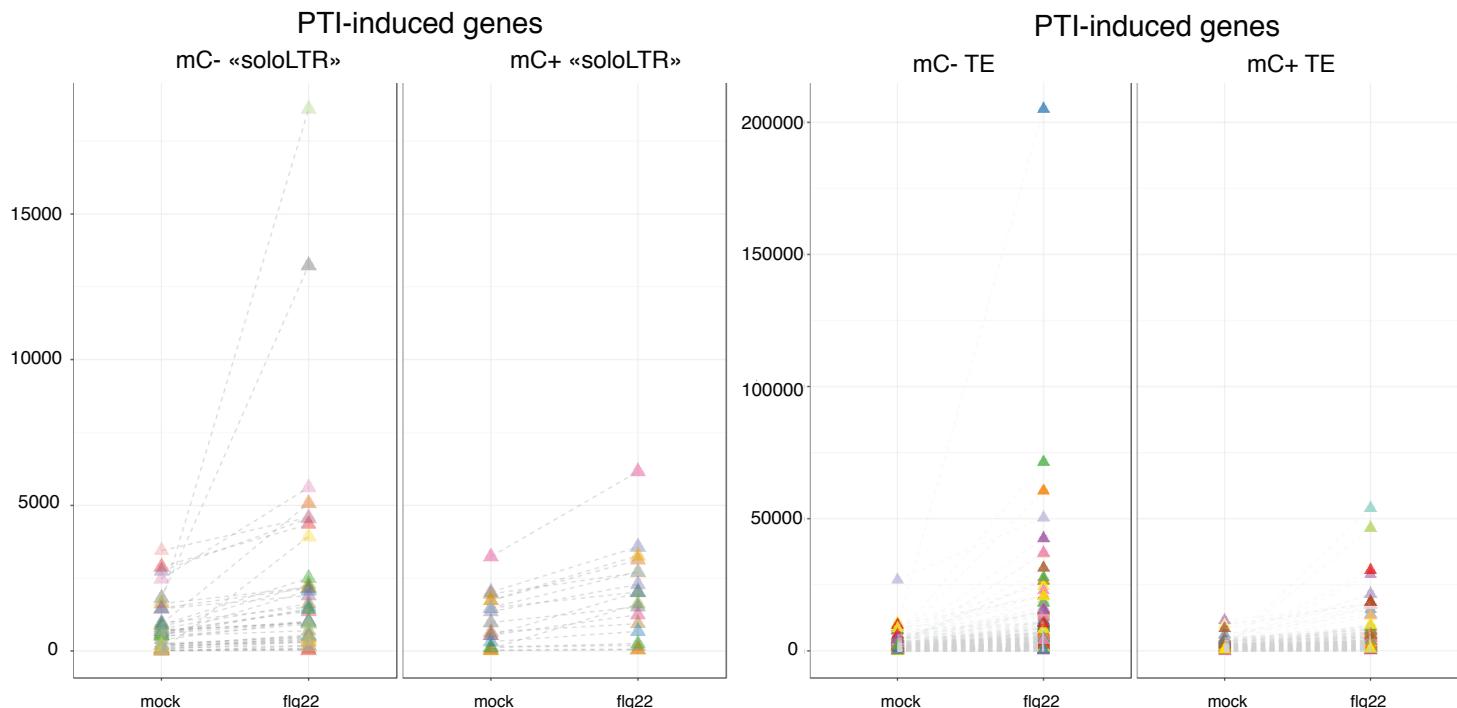
A**B****C**

Figure S2. Genome-wide analysis of the presence and methylation status of «soloLTRs» (LTR-containing TEs<1kB) and TEs<1kB in the promoter of genes induced during innate immune response.

A. Percentage of immunity-induced genes which promoter overlaps with a soloLTR (left panel, “soloLTRs+ promoter”) or a TE<1kB (right panel, “TE+ promoter”).

B. TE methylation status breakdown: percentages of immunity-induced genes with an unmethylated or methylated soloLTR in their promoter (left panel, “mC- soloLTR promoter” or “mC+ soloLTR promoter”) and percentages of immunity-induced genes with an unmethylated or methylated TE (right panel, “mC- TE promoter” or “mC+ TE promoter”). A soloLTR or TE(<1kB) is considered unmethylated when the average levels of CG, CHG and CHH methylation are all lower than 30%, 20% and 10% respectively. n/a : no information for DNA methylation.

C. Gene expression levels (normalized total counts) of individual genes from A and B pannels that have either an unmethylated (mC-) or methylated (mC+) soloLTR (left panel)/TE<1kB (right panel) in their promoter and 6 hours post-treatment with water or flg22.

mC-soloLTR PTI-induced genes				mC+soloLTR PTI-induced genes				n/a soloLTR PTI-induced genes			
AGI	Mock	Flg22	Fold-change	AGI	Mock	Flg22	Fold-change	AGI	Mock	Flg22	Fold-change
AT1G13320	1474,394	2157,401	1,46324649	AT1G16670	531,4821	2017,464	3,7959202	AT4G0920	133,4512	253,9608	1,9030242
AT1G14170	926,9942	5068,32	5,46747715	AT1G27000	1724,649	3131,827	1,8159211	AT4G0925	90,97992	166,1346	1,8260583
AT1G20930	29,20998	61,29382	2,09838612	AT2G04740	730,8223	958,6263	1,3117091				
AT1G21240	102,6169	190,8859	1,8601798	AT2G16950	2019,849	3572,915	1,7689026				
AT1G28280	819,1542	2277,281	2,78003938	AT2G34840	544,2363	1241,132	2,2805013				
AT1G47603	15,78963	94,36098	5,97613545	AT3G01410	28,68972	58,20288	2,0287018				
AT1G50730	767,5984	1625,333	2,11742621	AT3G28715	1953,824	2696,576	1,3801529				
AT1G51680	1819,711	13238,52	7,27506999	AT3G43230	1474,745	2024,652	1,3728822				
AT1G56560	668,7999	1011,499	1,51240876	AT3G46385	126,0982	189,404	1,5020363				
AT1G65540	1627,721	2187,088	1,34364986	AT3G50190	19,04037	54,2604	2,8497551				
AT1G65820	2742,804	4553,981	1,66033771	AT3G52710	1349,143	2285,239	1,6938442				
AT1G65840	676,8169	1364,678	2,01631726	AT3G52730	3237,026	6157,837	1,9023132				
AT1G66900	590,3268	987,1532	1,67221483	AT4G09310	133,4512	253,9608	1,9030242				
AT1G67410	179,4052	298,3769	1,66314524	AT4G11380	1736,601	3273,304	1,8848917				
AT1G73730	519,2483	692,5906	1,33383324	AT4G11790	649,0576	936,7273	1,4432113				
AT1G73740	389,8622	1022,075	2,62163263	AT4G29790	977,7299	1513,664	1,5481417				
AT2G06510	203,6876	547,8765	2,68978771	AT5G35700	1484,952	2720,684	1,8321694				
AT2G28890	1720,677	18610,07	10,8155536	AT5G39760	321,617	672,5703	2,091215				
AT2G31060	888,8931	2500,863	2,81345721								
AT2G34480	3446,345	4590,555	1,3320067								
AT3G02850	10,64682	32,06113	3,01133452								
AT3G19280	223,5963	369,2262	1,65130701								
AT3G26670	526,2773	2074,744	3,94230159								
AT3G27997	143,5347	410,3969	2,85921751								
AT3G44400	549,7666	1432,131	2,60497957								
AT3G56070	259,0853	454,1837	1,75302778								
AT3G56190	2891,881	4349,884	1,50417111								
AT3G57740	58,63522	315,8437	5,3865867								
AT4G00660	963,5033	1514,024	1,57137356								
AT4G14220	626,3119	1433,884	2,2894087								
AT4G32210	1447,792	1902,166	1,31383887								
AT4G35985	72,33066	481,5386	6,65746234								
AT4G35987	163,2113	327,7378	2,0080575								
AT4G39230	212,0943	514,999	2,42816007								
AT5G01050	1,805968	42,64019	23,6107124								
AT5G07270	699,2489	965,1699	1,38029519								
AT5G20020	2481,148	5610,75	2,26135273								
AT5G35320	720,9096	1329,868	1,84470876								
AT5G42960	498,6471	922,0047	1,84901259								
AT5G45090	13,5495	3934,186	290,356499								
AT5G52020	1,444774	168,3012	116,489616								
Mean	784,8	2236	13,10453	Mean	1058	1876	1,911405				

Appendix Table S1.

Normalized read counts associated with the immunity-induced genes with an unmethylated or methylated soloLTR in their promoter ("mC- soloLTR promoter" or "mC+ soloLTR promoter") and plotted on the left panel of Figure S2C . Genes with no information for DNA methylation (n/a soloLTR PTI-induced genes) are also shown.

Supplementary methods

To perform the genome-wide analysis presented in Figure S1, we decided on the criteria below:

- To define a set of “Immunity-induced genes” (“PAMP-induced genes”), we used our unpublished RNA-seq data (strand-specific, single-end sequencing) generated from 5 week-old leaves collected at 6 hours post-infiltration with either water (mock) or 1 µM of flg22 peptide (derived from two independent experiments). Data are accessible on the NCBI sequence read archive (SRA) under accession SRP133028 and a complete analysis of this dataset will be published elsewhere. Briefly, after quality control, reads were mapped to the *Arabidopsis Thaliana* genome (Tair 10) using TopHat (v2.0.8b) allowing mismatches and multiple hits (in this case, the score was divided by the number of total hits). The read counts were normalized to scale the raw library sizes using trimmed mean of M-values (TMM) proposed by Robinson and Oshlack (2010). Bioconductor package edgeR 3.0.8 (Robinson et al., 2010) was used to perform a differential expression analysis between mock- and flg22-treated datasets. The likelihood ratio test was used to compute the p-values for differential expression ($P<0.05$) (Benjamini and Hochberg, 1995) with a \log_2 fold-change >0 in “flg22” *versus* “mock” samples. We obtained a list of 6161 “PAMP-induced genes” (PCGs, pseudogenes, miRNA and other RNAs genes).
- We defined by “promoter” the sequence between the Transcriptional Start Site and 1kb upstream of the TSS and considered that a gene contains a TE in its promoter when a TE-overlapped with that 1kB sequence.
- There is currently no genome-wide annotation of soloLTRs (*i.e.*, rigorously, the product of homologous recombination between the two LTRs of a retroelement) in *Arabidopsis*. We thus retrieved “single LTRs” as a proxy for soloLTRs or derived-sequences thereof as they potentially have a *cis*-regulatory function and refer to them as “soloLTRs”. We defined by “soloLTR” a piece of LTR-retrotransposon which is less than 1kb and has homology with a LTR (by blast against the RepBase TE LTR-retrotransposons database). For the “TE” control group, current annotation from TAIR10 was used and TEs <1 kB selected.
- To determine the methylation status of each TE/LTR in the promoter of “immunity-induced genes”, we averaged the methylation levels for each context of methylation: CG, CHG and CHH using a published methylome (Kawakatsu *et al.*, 2016). A LTR or

TE was considered unmethylated when the average levels of CG, CHG and CHH methylation were all lower than 30%, 20% and 10% respectively. If the average methylation levels of either CG, CHG or CHH was higher than 30%, 20%, and 10% respectively, the TE/LTR was considered methylated. Genes that had in their promoter a LTR/TE with no methylation information were referred to as “n/a LTR/TE” and were not included in the gene expression analysis. Of note, *RPP4* was not recovered in our analysis as its induction in response to flg22 did not meet the p-value threshold of <0.05 (P=0.35) that we defined for “induced genes”.

References for supplementary methods

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289-300. Retrieved from <http://www.jstor.org/stable/2346101>

Kawakatsu T, Huang S shan C, Jupe F, Sasaki E, Schmitz RJJ, Urich MAA, Castanon R, Nery JRR, Barragan C, He Y, Chen H, Dubin M, Lee CR, Wang C, Bemm F, Becker C, O'Neil R, O'Malley RCC, Quarless DXX, Alonso-Blanco C, et al (2016) Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell* **166**: 492–506

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), R25. <http://doi.org/10.1186/gb-2010-11-3-r25>

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <http://doi.org/10.1093/bioinformatics/btp616>

Primer Name	Gene Name	Sequence
Bisulfite		
P365 GUS-F-Bs	<i>LTR-EVD</i> transgene	AAATATCACCACCTTATACAAAAAAACTAAAT
P366 <i>EVD-5'LTR</i> R2	<i>LTR-EVD</i> transgene	GTGYGAYAAGATYGAATGTAGTTATTG
P367 <i>EVD-5'LTR</i> F	<i>LTR-EVD</i> endogene	ACATRATCTTATRCTCTRATACCAT
P368 <i>EVD-5'LTR</i> R1	<i>LTR-EVD</i> endogene	GTAGAGGAYAAATGTTAATTYGTGTTGG
Chop-Assay(Sau96I)		
P342 <i>LTR</i> -GUStr (attB2rev)	<i>LTR-EVD</i> transgene	ACCACTTTGTACAAGAAAGCTGGGT
P343 <i>LTR-EVD</i> -tr	<i>LTR-EVD</i> transgene	CTCTATCTAACAGTCGCGACAAGATCG
P155 <i>EVD</i> sau96I-F	<i>LTR-EVD</i> endogene (fig 1)	CCCTAGAACACGGATTGGCA
P156 <i>EVD</i> sau96I-R	<i>LTR-EVD</i> endogene (fig 1)	TCGTGAGTCCTCTAACGG
P3F(P437) <i>EVD LTR</i>	<i>LTR-EVD</i> endogene (fig 3)	TCGTTGTTGGTCGATGTCATC
P522 EVDchopII R	<i>LTR-EVD</i> endogene (fig 3)	TCGTTGTTGGTCGATGTCATC
RT- qPCR		
UBI F	<i>UBIQUITIN</i>	TGAAGTCGTGAGACAGCGTTG
UBI R	<i>UBIQUITIN</i>	GGGCTTCTCATTGTTGGTC
WRKY29 F	<i>WRKY29</i>	TATCCAACGGATCAAGAGCTGA
WRKY29 R	<i>WRKY29</i>	TTTCTTGCCAAACACCCTTT
GUS F	<i>GUS</i>	ACCTCTCTTAGGCATTGGTTCG
GUS R	<i>GUS</i>	TCACCGCGCTATCAGCTTTAAC
P362 COPIA93 F	<i>COPIA93</i>	CGCTTGACCCGAAGATTCTC
P363 COPIA93 R	<i>COPIA93</i>	TTTCCACCAATAGCCCCGG
P519 <i>RPP4</i> 3' F	<i>RPP4</i>	ATGGTCGCCGACTTACAGAC
P520 <i>RPP4</i> 3'R	<i>RPP4</i>	TAC CCA ATATAC GTC GCC ATC C
ChIP		
P543 <i>soloLTR</i> -RPP4	<i>soloLTR-5/IGR</i> (RPP4)	TTTGTGGTTATCAGTTGTCTTG
P544 <i>soloLTR</i> - RPP4	<i>soloLTR-5/IGR</i> (RPP4)	TTCTCACTCATATACCAACTCTCTTG
P537 <i>RPP4</i> utrDISTALbis	<i>soloLTR-5</i> (RPP4)	TAAAGAAAATCGGCCCTTC
P538 <i>RPP4</i> utrDISTALbis	<i>soloLTR-5</i> (RPP4)	TGCTACAATTCCGCATATTCTT
P3F(P437) <i>EVD LTR</i>	<i>LTR-EVD</i> endogene	TCGTTGTTGGTCGATGTCATC
P3R(P438) <i>EVD LTR</i>	<i>LTR-EVD</i> endogene	TCGGCCCACTCTTGTAG
P4F(P439) <i>ATR LTR</i>	<i>LTR-ATR</i> endogene	TGCATAAGTCTGCGCTTGA
P4R(P441) <i>ATR LTR</i>	<i>LTR-ATR</i> endogene	AGTCCTCTCAACGGCTACA
P9F (P465) <i>CDS EVD/LTR</i>	<i>LTR-EVD/ATR</i> endogene	GGTGATACTTCAGGGGGAAA
P9R(P466) <i>CDS EVD/LTR</i>	<i>LTR-EVD/ATR</i> endogene	TTTTCTTCCTCGGGAGATATAGT
p583 <i>TA3</i> F	<i>TA3</i>	TAGGGTTCTAGTTGATCTGTATTGAGCTC
P584 <i>TA3</i> R	<i>TA3</i>	TTGCTCTCAAATCTCAATTGAAGTTT
P519 <i>AT5G17120</i> F	<i>AT5G17120</i>	TGGTCTGCAGAGAAATAGGGA
P520< i>AT5G17120 R	<i>AT5G17120</i>	GGAACGACTGTGGAATTGGT
P557 <i>flc</i> F	<i>FLOWERING LOCUS C</i>	TTGCTTGATTAATTGGGGTTT
P558 <i>flc</i> R	<i>FLOWERING LOCUS C</i>	GCCGGTCTTCCATTGTAA
sequencing		
p179 <i>LTR-EVD</i> wbox	<i>LTR-EVD</i> wbox	CACCACTCTATCTAACAGTCGCGACAAGATCGATG
p360 <i>soloLTR-5</i> (RPP4 wbo)	<i>soloLTR-5</i> (RPP4) wbox	AAAGACACACAAAGGAATATGAACGTTACAATTAAACGTAAGAC
genotyping		
MEF2 <i>met1</i> wt F	<i>MET1</i>	GCCTGGTCAAGTGGACTTCAT
MER2 <i>met1</i> wt R	<i>MET1</i>	CCATTCTTCACAGAGCATGCC
MEF1 <i>met1</i> mut F	<i>MET1</i>	GATTGTGTCCTACTACAGAGGC
TL2 <i>met1</i> mut R	<i>MET1</i>	TGGACGTGAATGTAGACACGTCG
P399 <i>ddm1</i> F	<i>DDM1</i>	ACGAAGCAACCAAGGAAGAA
P400 <i>ddm1</i> R	<i>DDM1</i>	GAGCCATGGGTTGTGAAACGTA
P520 <i>clf</i> LP	<i>CLF</i>	TCGACCCACTACAGACTGGTC
P530 <i>clf</i> RP	<i>CLF</i>	TTTGGGTTCTTTAGGAACC
Lba1.3	salk T-DNA	ATTTGCCGATTCGGAA
ecEVD		
P275 genewalkadaptator 1	adaptator	GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCGGCTGGT
P276 genewalkadaptator 2	adaptator	ACCAAGCCC
P278 genewalk AP2	adaptator (PCR/qPCR)	ACTATAGGGCACGCGTGGT
P279 <i>TY1</i> 5184 F	<i>LTR-EVD</i>	CACAAAGGTTACTATCTAATTCTATCAAT
P277 genewalk AP1	adaptator (PCR)	GTAATACGACTCACTATAGGGC
pyrosequencing		
P465 EVDATRChIP2F	<i>LTR- EVD/ATR</i>	GGTGATACTTCAGGGGGAAA
P466 EVDATRChIP2R	<i>LTR-EVD/ATR</i>	TTTTCTTCCTCGGGAGATATAGT
P491 seqPrevPyro	<i>LTR-EVD/ATR</i>	CTCCGTTCTTCTTCTTCTT

Appendix Table S2. Primers used in the study